working paper
department
of economics

STOCHASTIC EQUICONTINUITY AND WEAK CONVERGENCE
OF UNBOUNDED SEQUENTIAL EMPIRICAL PROCESES

Jushan Bai

94-7 Jan. 1993

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

STOCHASTIC EQUICONTINUITY AND WEAK CONVERGENCE
OF UNBOUNDED SEQUENTIAL EMPIRICAL PROCESES

Jushan Bai

94-7 Jan. 1993

Abstract

This note explains why sequential empirical processes arise naturally in the context of structural change and then provides an elementary proof for their stochastic equicontinuity. An application is considered for testing structural change in a linear regression and in a single equation of a simultaneous equations system.

Key words and phrases: Stochastic equicontinuity, weak convergence, sequential empirical process, two-parameter Brownian bridge, structural change.

Stochastic Equicontinuity and Weak Convergence of Unbounded Sequential Empirical Processes

Jushan Bai
Massachusetts Institute of Technology
(617) 253-6217
email: jbai@athena.mit.edu

Revised: January 1994

1 Introduction

A number of authors have recently studied stochastic equicontinuity for unbounded empirical processes for heterogeneous and dependent observations, e.g., Andrews [2], DeJong [11], and Hansen [16]. A review is given by Andrews [3]. In this paper we first introduce a weighted sequential empirical process and then study its stochastic equicontinuity. A sequential empirical process involves partial sums of stochastic processes. As such, an empirical process in the usual sense may be viewed as a special case of a sequential empirical process. Andrews [2] shows how empirical process theory can be used in various applications in econometrics, particularly in establishing asymptotic properties of econometric estimators and test statistics. We show in this paper that sequential empirical processes arise naturally in the context of structural change. Therefore they will be useful for testing parameter constancy in econometric models.

Sequential empirical processes and their related stochastic equicontinuity are first discussed by Bickel and Wichura [9], but with no particular motivation. In this paper, we consider a weighted sequential empirical process. The consideration of a weighted version is motivated by structural change in linear regressions, as explained in the next section. The weights are the regression variables or a set of instrument variables, so that the process involves unbounded summands in contrast to the one considered by Bickel and Wichura. Also regression variables are generally stochastic and serially correlated (e.g., autoregressive models); thus we have a stochastically weighted and dependent empirical process. This kind of sequential empirical process is not discussed much in the literature. Its stochastic equicontinuity may not be directly derived from the existing results. In addition, the contemporaneous treatment of dependent empirical processes demands highly technical and abstract analysis. In this paper, we provide an elementary proof for stochastic equicontinuity. The argument is a direct extension of that of Billingsley [10]. Because of the concrete structure of the process, we are able to derive concrete sufficient conditions for its stochastic equicontinuity. In

addition, an elementary proof is instructive.

Application of the result to structural changes in linear regressions is briefly discussed. The analysis is applicable to changes in a single equation of a simultaneous equations system. Tests based on a weighted sequential empirical process can detect changes in regression parameters and changes in variance. Most importantly, these tests can detect changes in error distribution functions that are not necessarily manifested in the form of changing variances. In other words, the test is able to test changes in higher moments of the data, whereas the CUSUM, fluctuation, and Wald type of tests may not be able to.

2 Sequential empirical process

For a given sequence of random variables $Z_1, Z_2, ..., Z_n$, the sequential empirical process of this sequence is defined as

$$B_n(k,z) = \frac{1}{\sqrt{n}} \sum_{t=1}^{k} \{I(Z_t \le z) - F(z)\}, \quad (k = 1, 2, ..., n)$$

where $F(z)$ is the distribution function of Z_i. Bickel and Wichura [9] first introduce and establish the stochastic equicontinuity of B_n for i.i.d. Z_i's. This is a two-parameter process. The summand in the process is bounded by 1. We shall consider a weighted sequential empirical process which is not bounded and may be dependent. Weighted sequential empirical processes arise naturally in the context of structural change. Consider the following linear regression model:

$$y_t = x_t'\beta + \varepsilon_t \quad (t = 1, 2, ..., n) \tag{1}$$

where x_t is a vector of explanatory variables, β is an unknown vector, and ε_t are i.i.d. disturbances with a continuous distribution function. Define the vector process:

$$S_n(z) = \sum_{t=1}^{n} x_t I(y_t \le z), \qquad -\infty < z < \infty.$$

In terms of parameter inference, observing this process is equivalent to observing the whole data set, with probability one. The vector process simply orders the original

data set according to the magnitude of the dependent variable. In this sense, $S_n(z)$ is a sufficient statistic for β (may be a "sufficient process" is more appropriate). Now suppose the true model obeys a two-regime regression:

$$y_t = x_t'\beta_1 + \varepsilon_t \quad (t = 1, 2, ..., n_1) \tag{2a}$$

$$y_t = x_t'\beta_2 + \varepsilon_t \quad (t = n_1 + 1, ..., n). \tag{2b}$$

To estimate β_1 and β_2, it is important to know to which regime a given observation belongs. The process $S_n(z)$ does not convey this information. It is thus not a sufficient statistic (process) for β_1 and β_2. However,

$$S_n^1(z) = \sum_{t=1}^{n_1} x_t I(y_t \leq z)$$

and

$$S_n^2(z) = \sum_{t=n_1+1}^{n} x_t I(y_t \leq z)$$

are jointly sufficient for β_1 and β_2, because S_n^1 only orders the first n_1 observations within themselves, and similarly S_n^2 orders the last $n - n_1$ observations within themselves. However, when the regime-switching point n_1 is unknown, then (S_n^1, S_n^2) is no longer a sufficient statistic. In this case, sufficient statistics are given by all pairs of (S_n^1, S_n^2) for $n_1 = 1, 2, .., n$. Introduce

$$S_n(k, z) = \sum_{t=1}^{k} x_t I(y_t \leq z),$$

which is a sequential empirical process up to normalization and centering. For $k = n_1$, the process $S_n(k, z)$ is simply $S_n^1(z)$ and $S_n(n, z) - S_n(n_1, z)$ is simply $S_n^2(z)$. Thus, when n_1 is unknown, sufficient statistics are given by $S_n(k, z),\ k = 1, ..., n$.

As an example of a dependent sequential empirical process, consider a time series regression:

$$y_t = \mu + \rho_1 y_{t-1} + \cdots \rho_p y_{t-p} + z_t'\delta + \varepsilon_t. \tag{3}$$

Denote $x_t = (1, y_{t-1}, ..., y_{t-p}, z_t')'$. The sequential empirical process is defined as before. But x_t is serially correlated, yielding a dependent sequential empirical process.

The above process $S_n(k,z)$ only describes the data, it does not incorporate the model. To do this, we modify the process to

$$S_n(k,z,\gamma) = \sum_{t=1}^{k} x_t I(y_t \leq z + x_t'\gamma).$$

A linear structure is introduced in the above process. Also note, under model (1),

$$S_n(k,z,\beta) = \sum_{t=1}^{k} x_t I(\varepsilon_t \leq z).$$

However, the parameter β is unknown, so $S_n(k,z,\beta)$ is not observable or computable. To solve this problem one can replace β by an estimator, $\hat{\beta}$. If we put $\hat{\varepsilon}_t = y_t - x_t'\hat{\beta}$, then

$$S_n(k,z,\hat{\beta}) = \sum_{t=1}^{k} x_t I(\hat{\varepsilon}_t \leq z)$$

which can be considered an estimated sequential empirical process. This process embodies the model and data. The estimated parameter can be obtained from the first k observations or from the whole sample. In the former case, a sequence of estimators is needed, which may be obtained by recursive estimation. In our application, we use a whole-sample estimator. The test for parameter constancy is based on the process $S_n(k,z,\hat{\beta})$, see Section 4. Tests based on the weighted sequential empirical process are more powerful than those based on a non-weighted process, as pointed out by Bai [8]. To study the asymptotic property of the test, we need the weak convergence of $S_n(k,z,\hat{\beta})$, whose convergence in turn depends on the weak convergence of $S_n(k,z,\beta)$. The weak convergence of the latter is our focus. By normalizing and centering of S_n, we define the vector process

$$H_n(s,z) = (X'X)^{-1/2} \sum_{t=1}^{[ns]} x_t \{I(\varepsilon_t \leq z) - F(z)\} \tag{4}$$

where $X = (x_1, x_2, ..., x_n)'$. In the next section, we study the stochastic equicontinuity of H_n and its weak convergence.

The weighting vector x_t does not have to be the regression variables. Generally, it can be a set of instrumental variables. Consider a single equation in a simultaneous

equations system:

$$y_t = z_t'\beta + \varepsilon_t \tag{5}$$

where z_t includes other endogenous variables so that ε_t is correlated with z_t. If one uses z_t in place of x_t in the definition of H_n, then the process will not have a proper limit because the summands do not have zero mean. Now suppose x_t is a vector of instruments that is correlated with z_t but independent of ε_t. Then we can still consider the weak convergence of (4). We may call this process the instrumental-variable weighted sequential empirical process. Tests based on a instrumental-variable weighted process will have nontrivial local power only if X is a set of valid instruments in the sense that $\text{plim}(X'X/n)$ and $\text{plim}(X'Z/n)$ have full column rank and X is uncorrelated with ε_t, where $Z = (z_1', ..., z_n')'$.

3 Stochastic Equicontinuity

To derive the stochastic equicontinuity for weighted sequential empirical processes, we impose the following conditions (their implications are discussed below).

(A.1) The random variables ε_t are i.i.d. with a continuous distribution function F.

(A.2) The disturbances ε_t are independent of all contemporaneous and past regressors.

(A.3) The regressors $\{x_t\}$ form a triangular array (for simplicity the dependence on n is suppressed) and satisfy;

$$\text{plim}\frac{1}{n}\sum_{t=1}^{[ns]} x_t x_t' = Q(s) \quad \text{for } s \in [0,1],$$

where $Q(s)$ is a $p \times p$ nonrandom positive definite matrix for $s > 0$ and $Q(0) = 0$. The convergence is necessarily uniform in s, because the sum is "monotonic" in s.

(A.4)

$$\max_{1 \le t \le n} n^{-1/2}\|x_t\| = o_p(1).$$

(A.5) For every fixed s_1, there exists a random variable Z_n (may depend on s_1) such that, for all $s \ge s_1$,

$$\frac{1}{n}\sum_{t=[ns_1]}^{[ns]} \|x_t\| \le (s - s_1)Z_n$$

with probability one. In addition, the tail probability of Z_n satisfies, for some $\rho > 0$ and $M < \infty$:

$$P(|Z_n| > C) \leq M/C^{2(1+\rho)}.$$

(A.6) There exist $\gamma > 1$, $\alpha > 1$ and $K < \infty$ such that for all $0 \leq u \leq v \leq 1$, and for all n,

$$\frac{1}{n} \sum_{i<t\leq j} E(x_t'x_t)^\gamma \leq K(v-u) \quad \text{and} \quad E(\frac{1}{n} \sum_{i<t\leq j} x_t'x_t)^\gamma \leq K(v-u)^\alpha,$$

where $i = [nu]$, $j = [nv]$.

Assumption (A.2) allows for dynamic models, e.g., autoregressive model (3). Assumption (A.3) allows for trending regressors written in the form $x_t = g(t/n)$, for some function g. This assumption is often maintained in recursive estimation for constructing CUSUM tests, see, e.g., Ploberger, Kramer, and Kontrus [18]. Assumption (A.4) is conventional for linear models and is used for obtaining normality. Assumptions (A.5) and (A.6) are unique for our problem. They are the main assumptions for the equicontinuity of the sequential empirical process H_n. In (A.5), Z_n may be taken to be $\max_k k^{-1} \sum_{t=i}^{i+k} \|x_t\|$ provided the condition on the tail probability is also satisfied, where $i = [ns_1]$ is fixed. It is generally impossible, however, to choose $Z_n = O_p(1)$ uniformly in both s and s_1. If $E(x_t'x_t)^2 \leq M$ for all t, then (A.6) is satisfied with $\gamma = 2$ and $\alpha = 2$, because $E(\sum_{t=i}^{j} x_t'x_t)^2 \leq \{\sum_{t=i}^{j}[E(x_t'x_t)^2]^{1/2}\}^2$ by the Cauchy-Schwarz inequality. Finally, when regressors x_t are bounded (A.4)-(A.6) will be satisfied.

Let $\mathcal{T} = [0,1] \times \mathcal{R}$ be the parameter set with metric $\rho(\{r,y\},\{s,z\}) = |s - r| + |F(z) - F(y)|$. Let $D[\mathcal{T}]$ be the set of functions defined on $\mathcal{T}$ that are right continuous and have left limits. We equip $D[\mathcal{T}]$ with the Skorohod metric (Pollard [19]). The vector process H_n belongs to the Cartesian product space $D[\mathcal{T}]^p$, equipped with the corresponding product Skorohod topology. The weak convergence of H_n in the space $D[\mathcal{T}]^p$ is implied by the finite dimensional convergence together with stochastic equicontinuity. The latter condition also implies the sample path of the limiting process of H_n will be continuous with probability one.

Theorem 1 *Under assumptions (A.1), (A.2), (A.5), and (A.6), the process H_n is stochastically equicontinuous on $(\mathcal{T}, \rho)$. That is for any $\epsilon > 0$, $\eta > 0$, there exists a $\delta > 0$ such that for large n,*

$$P\left(\sup_{[\delta]} \|H_n(r,y) - H_n(s,z)\| > \eta\right) < \epsilon$$

where $[\delta] = \{(\tau_1, \tau_2); \tau_1 = (r,y), \tau_2 = (s,z), \rho(\tau_1,\tau_2) < \delta\}$ with $[\delta] \subset \mathcal{T} \times \mathcal{T}$.

When $x_t = 1$ for all t, the equicontinuity of H_n is implied by the result of Bickel and Wichura [9]. When $\{(x_t, \varepsilon_t)\}$ are independent, equicontinuity can also be proved by extending the method of Bickel and Wichura. It is the statistical dependence in data that requires a different framework of proof. Dependence in data could be a big obstacle for proving equicontinuity. Indeed, the powerful tool of symmetrization depends heavily on the independence assumption, although it is extended to m-dependent processes by Andrews [2]. Recent development explores ways of getting around the difficulty. And there are many successful results; examples are Andrews [1] for a smoothed class of functions of near-epoch variables, DeJong [11] for unbounded strong mixing processes, Doukhan, Massart and Rio [12] for unbounded absolutely regular processes, and Hansen [16] for unbounded mixingales. A review is provided by Andrews [3], also see Andrews and Pollard [5] for a bounded strong mixing sequence. For the proposed weighted sequential empirical process (4), the summands are unbounded martingale differences (for each fixed z). It seems that the method of Levental [17] may be used. The conditions of Levental, however, are not primitive and only work well for bounded martingales, though it is noted by Hansen [16] that Levental's method may be extended to unbounded ones. Hansen's own approach, as pointed out by the author himself, does not work well for indicator types of functions. We shall offer an elementary proof. Since our purpose is not to cover as wide a range of processes as possible, our conditions are specific and primitive. Given the concrete structure of our process, we feel an elementary argument is more instructive. We are also interested in the limiting process.

The proof is provided in the appendix. In the proof, we focus on the vector process

$$Y_n(s,u) = n^{-1/2} \sum_{t=1}^{[ns]} x_t \{I(U_t \le u) - u\} \tag{6}$$

where $U_1, U_2, ..., U_n$ are i.i.d. uniform on $[0,1]$ with U_t independent of x_j for $j \le t$. Effectively, we replace ε_t by $F(\varepsilon_t)$ which is uniform on $[0,1]$. So that $H_n(s,z) = (X'X/n)^{-1/2} Y_n(s, F(z))$. By assumption, $(X'X/n) \xrightarrow{p} Q(1)$, a positive definite matrix, so Y_n and H_n are equivalent in terms of stochastic equicontinuity.

Corollary 1 *Under assumptions (A.1)-(A.6), the process H_n converges weakly to a Gaussian process H with zero mean and covariance matrix*

$$E\{H(r,y)H(s,z)'\} = Q(1)^{-1/2} Q(r \wedge s) Q(1)^{-1/2} [F(z \wedge y) - F(z)F(y)]. \tag{7}$$

Proof. The finite dimensional convergence to a normal distribution follows from CLT for martingale differences. This together with Theorem 1 implies that H_n converges weakly to some process H. To verify the covariance matrix, consider the expected value of Y_n, for $r < s$ and $u = F(z) < v = F(y)$. Using double expectation and martingale property, we obtain

$$E\{Y_n(r,u) Y_n'(s,v)\} = \frac{1}{n} E\left(\sum_{t=1}^{[nr]} x_t x_t' \right) (u - uv) \tag{8}$$

which tends to $Q(r)(u - uv)$. From $(X'X/n)^{-1/2} \xrightarrow{p} Q(1)^{-1}$, we arrive at (7). □

We now introduce a Brownian bridge type process which is closely related to tests for parameter constancy in linear regressions. Let $X_k = (x_1, ..., x_k)'$, $X = (x_1, x_2, ..., x_n)'$, and

$$A_k = (X'X)^{-1/2} (X_k' X_k) (X'X)^{-1/2}. \tag{9}$$

The matrix $A_{[ns]}$ converges to $A(s) = Q(1)^{-1/2} Q(s) Q(1)^{-1/2}$. In the special case that $Q(s) = sQ$ for some positive definite matrix Q, $A(s) = sI$, where I is a $p \times p$ identity matrix.

Corollary 2 *Under the assumptions of Corollary 1, the process V_n defined as*

$$V_n(s,z) = H_n(s,z) - A_{[ns]}H_n(1,z)$$

converges weakly to a Gaussian process V with mean zero and covariance matrix

$$E\{V(r,y)V(s,z)'\} = \{A(r \wedge s) - A(r)A(s)\}\{F(y \wedge z) - F(y)F(z)\}. \tag{10}$$

Proof. The stochastic equicontinuity of V_n follows from that of H_n and the convergence of $A_{[ns]}$ to a deterministic matrix $A(s)$ uniformly in s. The limiting process of V_n is, by Corollary 1,

$$V(s,z) = H(s,z) - A(s)H(1,z).$$

Now (10) follows easily from (7). □

As noted earlier, when $Q(s) = sQ$ for some $Q > 0$, $A(s)$ becomes sI and the covariance matrix of V becomes $(r \wedge s - rs)\{F(z \wedge y) - F(z)F(y)\}I$. A process $B(s,u)$ is said to be a two-parameter Brownian bridge on $[0,1]^2$ if it is a zero-mean Gaussian process with covariance function

$$EB(r,u)B(s,v) = (r \wedge s - rs)(u \wedge v - uv).$$

We see that $V(s,z)$ has the same distribution as $B^*(s, F(z))$, where B^* is a vector of p independent Brownian bridges.

4 An Application in Structural Change

Consider the structural change model (2). The objective is to test the null hypothesis $H_0 : \beta_1 = \beta_2$ with n_1 unknown. There is a rich literature on the problem, for example, Andrews [4]. Here we construct a test using an estimated sequential empirical process. We estimate model (1) by OLS or other methods and compute the residuals by $\hat{\varepsilon}_t = y_t - x_t'\hat{\beta}$. Define the $p \times 1$ vector process T_n,

$$T_n(\frac{k}{n}, z) = (X'X)^{-1/2} \sum_{t=1}^{k} x_t I(\hat{\varepsilon}_t \le z) - A_k (X'X)^{-1/2} \sum_{t=1}^{n} x_t I(\hat{\varepsilon}_t \le z) \tag{11}$$

and the test statistic

$$M_n = \max_k \sup_z \|T_n(\frac{k}{n}, z)\|_\infty$$

where $\|y\|_\infty = \max\{|y_1|, ..., |y_p|\}$, the maximum norm. The process T_n takes at most n^2 different values, so the maximum value always exists. The actual computation of M_n is straightforward. If x_t contains a constant regressor (we do assume this), then

$$(X'X)^{-1/2} \sum_{t=1}^{k} x_t - A_k(X'X)^{-1/2} \sum_{t=1}^{n} x_t \equiv 0 \quad \text{for all } k$$

so that $I(\hat{\varepsilon}_t \leq z)$ can be replaced by $I(\hat{\varepsilon}_t \leq z) - F(z)$ without changing the value of T_n. Therefore, T_n is centered (only approximately centered because F is not the d.f. of $\hat{\varepsilon}_t$). Recognizing this, we see that T_n is the same as V_n defined in Corollary 2 except T_n uses estimated residuals while V_n uses true disturbances.

Using the result of this paper, Bai [7] shows that if the residuals are obtained from a root-n consistent estimator of β, then

$$T_n(\frac{[ns]}{n}, z) \Rightarrow B^*(s, F(z))$$

where $B^* = (B_1, B_2, ..., B_p)$ is a vector of p independent two-parameter Brownian bridges defined on $[0,1]^2$. A similar result is obtained by Bai [6] for bounded (non-weighted) sequential empirical processes based on ARMA residuals. By the continuous mapping theorem,

$$M_n \xrightarrow{d} \max_{0 \leq s,u \leq 1} \|B^*(s,u)\|_\infty.$$

The test M_n is asymptotically distribution free. Critical values are tabulated in Bai [8]. It is interesting to realize that the limiting process T_n does not depend on the estimated parameters. The underlying reason is that the process T_n consists of two terms. The estimation effects are canceled out in the first term and the second term. This is in contrast with the classical goodness-of-fit test where the estimation effect does not go away, see Durbin [13] and [14].

As for the limiting distribution under local alternatives, we consider a single equation in a simultaneous equations system. The model under the null hypothesis is given

by (5). Suppose the alternative hypothesis postulates that

$$y_t = z_t'\beta_t + \varepsilon_t \tag{12}$$

where $\beta_t = \beta[1 + \frac{1}{\sqrt{n}}g(t/n)]$ with g a vector-valued function defined on $[0,1]$ and Riemann-Stieltjes integrable. Suppose x_t is a vector of instrumental variables. For simplicity, assume in (A.3) $Q(s) = sQ_{xx}$ for some $Q_{xx} > 0$. Also assume $\text{plim}\frac{1}{n}\sum_{t=1}^{[ns]} x_t z_t' = sQ_{xz}$, a $p \times r$ matrix. Then Bai [8] shows that

$$M_n \xrightarrow{d} \max_{0 \le s,u \le 1} \|B^*(s,u) + p(u)Q_{xx}^{-1/2}Q_{xz}G(s)\|_\infty$$

where $p(u) = f(F^{-1}(u))$ and $G(s) = \int_o^s g(v)dv - s\int_0^1 g(v)dv$. Note that $G(s) \equiv 0$ if and only if g is a constant vector, implying no change in β_t. In order for the test to have non-trivial local power for "all" non-constant $g's$, Q_{xz} must have a full column rank. Otherwise, there exists a non-zero $G(s)$ such that $Q_{xz}G(s) \equiv 0$ so that M_n will have the same limiting distribution under both the null and alternative hypotheses. In summary, when valid instruments are available, M_n can be used to test changes in a simultaneous equations system and M_n possesses non-trivial local power. Regressors themselves constitute valid instruments when they are independent of error disturbances.

The same test can also detect changes in the following type:

$$y_t = x_t'\beta + \varepsilon_t^*$$

with ε_t^* having a continuous d.f. F for $t \le n_1$ and ε_t^* having a continuous d.f. G for $t > n_1$. Bai [8] argues that even if the two distributions have the same mean and same variance, as long as $F \ne G$, it is detectable by M_n, whereas the fluctuation, CUSUM and Wald tests may fail to diagnose this kind of shift.

A Appendix

Lemma A.1 *Assume the conditions of Theorem 1 hold. Then there exists a $K < \infty$, such that for all $s_1 < s_2$ and $u_1 < u_2$, where $0 \le s_i, u_i \le 1$ $(i = 1, 2)$*

$$E\|Y_n(s_2,u_2) - Y_n(s_1,u_2) - Y_n(s_2,u_1) + Y_n(s_1,u_1)\|^{2\gamma}$$
$$\le K(u_2-u_1)^\alpha(s_2-s_1)^\alpha + n^{-(\gamma-1)}K(u_2-u_1)(s_2-s_1).$$

Without the loss of generality, one can assume that $\alpha \le \gamma$, since $|u_2 - u_1| \le 1$ and $|s_2 - s_1| \le 1$. Moreover, when

$$\tau n^{-(\gamma-1)/2(\alpha-1)} \le u_2 - u_1 \quad \text{and} \quad \tau n^{-(\gamma-1)/2(\alpha-1)} \le s_2 - s_1 \tag{13}$$

for $\tau > 0$, the lemma implies

$$E\|Y_n(s_2,u_2) - Y_n(s_1,u_2) - Y_n(s_2,u_1) + Y_n(s_1,u_1)\|^{2\gamma}$$
$$\le K[1+\tau^{-2(\alpha-1)}](u_2-u_1)^\alpha(s_2-s_1)^\alpha. \tag{14}$$

This inequality is analogous to (22.15) of Billingsley ([10], p. 198).

Proof. Write $\eta_t = I(u_1 < U_t \le u_2) - u_2 + u_1$ and $Y_n^* = Y_n(s_2,u_2) - Y_n(s_1,u_2) - Y_n(s_2,u_1) + Y_n(s_1,u_1)$ for the moment. Then $Y_n^* = n^{-1/2}\sum_{i<t\le j} x_t\eta_t$ with $i = [ns_1]$ and $j = [ns_2]$. Note that $\{x_t\eta_t, \mathcal{F}_t\}$ is a sequence of (nonstationary) vector martingale differences, where $\mathcal{F}_t$ is the σ-field generated by $\ldots, x_t, x_{t+1}; \ldots, U_{t-1}, U_t$. By the inequality of Rosenthal (Hall and Heyde [15] p. 23), there exists a constant $M < \infty$ only depending on γ and p such that

$$E\|Y_n^*\|^{2\gamma} = E\left\{\left(\frac{1}{n}[\sum_{i<t\le j} x_t\eta_t]' \sum_{i<h\le j} x_h\eta_h\right)^\gamma\right\}$$
$$\le ME\left(\frac{1}{n}\sum_{i<t\le j} E\{(x_t'x_t)\eta_t^2|\mathcal{F}_{t-1}\}\right)^\gamma + Mn^{-\gamma}\sum_{i<t\le j} E\{(x_t'x_t)^\gamma\eta_t^{2\gamma}\}. \tag{15}$$

Note that x_t is measurable with respect to $\mathcal{F}_{t-1}$ and η_t is independent of $\mathcal{F}_{t-1}$. In addition, $E\eta_t^2 \le u_2 - u_1$ and $E\eta_t^{2\gamma} \le u_2 - u_1$. These results together with assumption

(A.6) provide bounds for the two terms on the right of (15). The first term is bounded by

$$M(u_2-u_1)^\gamma E\left(\frac{1}{n}\sum_{i<t\leq j}(x_t'x_t)\right)^\gamma \leq MK(u_2-u_1)^\gamma(s_2-s_1)^\alpha$$

and the second term is bounded by

$$Mn^{-(\gamma-1)}(u_2-u_1)\frac{1}{n}\sum_{i<t\leq j}E(x_t'x_t)^\gamma \leq MKn^{-(\gamma-1)}(u_2-u_1)(s_2-s_1).$$

Renaming MK as K, the lemma follows from $(u_2-u_1)^\gamma \leq (u_2-u_1)^\alpha$, for $\gamma \geq \alpha$.

Lemma A.2 *Under (A.5), we have for $s_1 \leq s \leq s_2$ and $u_1 \leq u \leq u_2$,*

$$\|Y_n(s,u)-Y_n(s_1,u_1)\| \leq \|Y_n(s_2,u_2)-Y_n(s_1,u_1)\| + O_p(1)n^{1/2}[(u_2-u_1)+(s_2-s_1)]$$

where the term $O_p(1)$ is uniform in s $(s \geq s_1)$, does not depend on u and u_1 and satisfies

$$P(|O_p(1)|>C) < M/C^{2(1+\rho)}, \quad \forall C>0, \quad \textit{for some } \rho>0.$$

Proof. First notice that all of the components of x_t can be assumed to be nonnegative. Otherwise write $x_t=\sum_{i=1}^p x_t^+(i)-\sum_{i=1}^p x_t^-(i)$ where $x_t^+(i)=(0,..0,x_{ti},0,...,0)'$ if $x_{ti}\geq 0$ and $x_t^-(i)=(0,..0,-x_{ti},0,...,0)'$ if $x_{ti}<0$. In this way, Y_n can be written as a linear combination (with coefficients 1 or -1) of at most $2p$ processes with each process having nonnegative weighting vectors. In addition, $\|x_t^+(i)\| \leq \|x_t\|$ and $\|x_t^-(i)\| \leq \|x_t\|$. So assumptions (A.5) and (A.6) are satisfied for $x_t^+(i)$ and $x_t^-(i)$. It is thus enough to assume that the x_t are nonnegative. A new piece of notation, for vectors a and b, take $a \leq b$ to mean $a_i \leq b_i$ for all components. Since $x_t \geq 0$, the vector functions $x_tI(U \leq u)$ and x_tu are nondecreasing in u. It is easy to show

$$\begin{aligned}Y_n(s,u)-Y_n(s_1,u_1) &\leq Y_n(s_2,u_2)-Y_n(s_1,u_1)\\ &+n^{1/2}(\frac{1}{n}\sum_{t=1}^{[ns]}x_t)(u_2-u)+n^{1/2}\left(\frac{1}{n}\sum_{t=[ns]}^{[ns_2]}x_t\{I(U_t\leq u_2)-u_2\}\right)\end{aligned}$$

and

$$Y_n(s_1,u_1)-Y_n(s,u) \leq n^{1/2}(\frac{1}{n}\sum_{t=1}^{[ns]}x_t)(u-u_1)+n^{1/2}\left(\frac{1}{n}\sum_{t=[ns_1]}^{[ns]}x_t\{I(U_t\leq u)-u_1\}\right).$$

The lemma follows from the boundedness of the indicator function and (A.5). $\square$

Proof of Theorem 1. We shall evaluate directly the modulus of continuity. Define

$$\omega_\delta(Y_n) = \sup\{\|Y_n(s',u') - Y_n(s'',u'')\|;\ |s'-s''| < \delta, |u'-u''| < \delta, s', s'', u', u'' \in [0,1]\}.$$

We shall show that for any $\epsilon > 0$ and $\eta > 0$, there exist a $\delta > 0$ and an integer n_0, such that

$$P(\omega_\delta(Y_n) > \epsilon) < \eta, \qquad n > n_0.$$

Since $[0,1]^2$ has only about δ^{-2} squares with side length δ, it suffices to show that for every point $(s_1, u_1) \in [0,1]^2$, every $\epsilon > 0$ and $\eta > 0$, there exist a $\delta \in (0,1)$ and an integer n_0 such that

$$P(\sup_{\langle\delta\rangle} \|Y_n(s,u) - Y_n(s_1,u_1)\| > 5\epsilon) < 2\delta^2\eta, \qquad n > n_0. \tag{16}$$

where $\langle\delta\rangle = \{(s,u);\ s_1 \le s \le s_1 + \delta, u_1 \le u \le u_1 + \delta\} \cap [0,1]^2$.

For a given $\delta > 0$ and $\eta > 0$, choose C large enough so the $O_p(1)$ in Lemma A.2 satisfies

$$P(|O_p(1)| > C) < \delta^2\eta. \tag{17}$$

By Lemma A.2 (see also (22.18) of Billingsley [10], p. 199), when $|O_p(1)| \le C$,

$$\sup_{\langle\delta\rangle} \|Y_n(s,u) - Y_n(s_1,u_1)\| \le 3 \max_{1\le i,j\le m} \|Y_n(s_1 + i\epsilon_n, u_1 + j\epsilon_n) - Y_n(s_1,u_1)\| + 2\epsilon$$

where $\epsilon_n = \epsilon/(n^{1/2}C)$ and $m = [n^{1/2}C\delta/\epsilon] + 1$. Write

$$X(i,j) = Y_n(s_1 + i\epsilon_n, u_1 + j\epsilon_n) - Y_n(s_1,u_1).$$

Then

$$P(\sup_{\langle\delta\rangle} \|Y_n(s,u) - Y_n(s_1,u_1)\| > 5\epsilon) \le P(|O_p(1)| > C) + P(\max_{1\le i,j\le m} \|X(i,j)\| > \epsilon). \tag{18}$$

Now for fixed i and k $(i \ge k)$ write $Z(j) = X(i,j) - X(k,j)$. Notice that

$$(\epsilon/C)n^{-(\gamma-1)/(\alpha-1)} \le \epsilon/(Cn^{1/2}) = \epsilon_n \le j\epsilon_n, \qquad j \ge 1,$$

which follows from $n^{-(\gamma-1)/2(\alpha-1)} \leq n^{-1/2}$ because $1 < \alpha \leq \gamma$. By (13) and (14),

$$E\|Z(j) - Z(l)\|^{2\gamma} \leq KC_\epsilon[(i-k)\epsilon_n]^\alpha[(j-l)\epsilon_n]^\alpha, \qquad 1 \leq l \leq j \leq m$$

where, from (14) with $\tau = \epsilon/C$,

$$C_\epsilon = [1 + (C/\epsilon)^{2(\alpha-1)}] \leq 2(C/\epsilon)^{2(\alpha-1)} \qquad \text{for small } \epsilon. \tag{19}$$

Thus by Theorem 12.2 of Billingsley ([10], p. 94), we have

$$P(\max_{1\leq j\leq m} \|Z(j)\| > \epsilon) \leq \frac{K_1 K C_\epsilon}{\epsilon^{2\gamma}}[(i-k)\epsilon_n]^\alpha (m\epsilon_n)^\alpha \leq \frac{K_2 C_\epsilon}{\epsilon^{2\gamma}}[(i-k)\epsilon_n]^\alpha \delta^\alpha \tag{20}$$

where K_1 is a generic constant and $K_2 = 2^\alpha K_1 K$. The last inequality follows from $(m\epsilon_n) \leq 2\delta$ for large n. Because

$$\left|\max_j \|X(i,j)\| - \max_j \|X(k,j)\|\right| \leq \max_j \|X(i,j) - X(k,j)\| = \max_j \|Z(j)\|,$$

if we let $V(i) = \max_j \|X(i,j)\|$, then (20) implies

$$P(|V(i) - V(k)| > \epsilon) < \frac{K_2 C_\epsilon}{\epsilon^{2\gamma}}[(i-k)\epsilon_n]^\alpha \delta^\alpha, \qquad 1 \leq k \leq i \leq m.$$

Thus by Theorem 12.2 of Billingsley once again [let $\xi_h = V(h) - V(h-1)$, so that $V(i)$ is the partial sum S_i of random variables ξ_h in Billingsley's notation], we obtain

$$P(\max_{1\leq i\leq m} |V(i)| > \epsilon) \leq \frac{K_1' K_2 C_\epsilon}{\epsilon^{2\gamma}}(m\epsilon_n)^\alpha \delta^\alpha \leq \frac{K_3 C_\epsilon}{\epsilon^{2\gamma}}\delta^{2\alpha}$$

where K_1' is a generic constant and $K_3 = 2^\alpha K_1' K_2$. Note that $\max_i |V(i)| = \max_i \max_j \|X(i,j)\|$. Thus by (18)

$$P(\sup_{(\delta)} \|Y_n(s,u) - Y_n(s_1,u_1)\| > 5\epsilon) \leq \delta^2\eta + \frac{K_3 C_\epsilon}{\epsilon^{2\gamma}}\delta^{2\alpha}.$$

By (19), the second term on the right hand side above is bounded by

$$\frac{K_3 C_\epsilon}{\epsilon^{2\gamma}}\delta^{2\alpha} \leq \delta^2 \frac{2K_3}{\epsilon^{2(\gamma+\alpha-1)}}(C\delta)^{2(\alpha-1)}. \tag{21}$$

By Lemma A.2, one can choose $C = (M/\eta)^{2(1+\rho)}\delta^{-(\frac{1}{1+\rho})}$ to assure (17) and the left hand side (21) becomes $K(\epsilon,\eta)\delta^a$, where $K(\epsilon,\eta)$ is a constant and $a = \frac{2(\alpha-1)\rho}{1+\rho} > 0$. By choose δ such that $K(\epsilon,\eta)\delta^a \leq \eta$, (16) follows. The proof of the theorem is completed.

□

References

[1] Andrews, D.W.K. An empirical process central limit theorem for dependent non-identically distributed random variables. *Journal of Multivariate Analysis* 38 (1991): 187-203.

[2] Andrews, D.W.K. Empirical process methods in econometrics. In *The Handbook of Econometrics*, Vol 4. Amsterdam: North-Holland, 1993.

[3] Andrews, D.W.K. An introduction to econometric applications of empirical process theory for dependent random random variables. *Econometric Reviews* 12 (1993): 183-216.

[4] Andrews, D.W.K. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61 (1993): 821-856.

[5] Andrews, D.W.K. and D. Pollard. A functional central limit theorem for strong mixing stochastic processes. Cowles foundation Discussion Paper No. 951, Cowles Foundation, Yale University, 1990.

[6] Bai, J. Weak convergence of sequential empirical processes of ARMA residuals. Forthcoming in the *Annals of Statistics*, 1991.

[7] Bai, J. On weighted sequential empirical process of regression residuals. Manuscript, Department of Economics, MIT, 1993.

[8] Bai, J. Testing for parameter constancy in linear regressions: an empirical distribution function approach. Manuscript, Department of Economics, MIT, 1993.

[9] Bickel, P.J. and M.J. Wichura. Convergence for multiparameter stochastic processes and some applications. *Annals of Mathematical Statistics* 42 (1971): 1656-1670.

[10] Billingsley, P. Convergence of Probability Measures. New York: Wiley, 1968.

[11] DeJong, R.M. Stochastic equicontinuity for unbounded mixing processes. Free University, Amsterdam, 1993.

[12] Doukhan, P., P. Massart and E.Rio. Invariance principles for absolutely regular empirical processes. University de Paris, 1992.

[13] Durbin, J. Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics* 1 (1973): 279-290.

[14] Durbin, J. Kolmogorov-Smirnov tests when parameters are estimated. In *Empirical Distributions and Processes*, Springer-Verlad Lecture Notes in Mathematics, 1976.

[15] Hall, P. and C.C. Heyde. Martingale Limit Theory and Its Application. New York: Academic Press, 1980.

[16] Hansen, B.E. Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays. Manuscript, Department of Economics, University of Rochester, 1993.

[17] Levental, S. A uniform CLT for uniformly bounded families of martingale differences. *Journal of Theoretical Probability* 2 (1988): 271-287.

[18] Ploberger, W., W. Kramer, and K. Kontrus. A new test for structural stability in the linear regression model. *Journal of Econometrics* 40 (1988): 307-318.

[19] Pollard, D. Convergence of Stochastic Processes. New York: Springer-Verlag, 1984.